ISSN: 2321-7782 (Online)
Volume 3, Issue 6, June 2015
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

A Survey on Taxonomy of Intrusion Detection System (IDS) and Anomaly Based Intrusion Detection System (ABIDS) Techniques

Meesala Shobha Rani[1]
PG Scholar,
Department of Computer Science & Engineering
Karunya University, Coimbatore, India

S. Basil Xavier[2]
Assistant Professor,
Department of Computer Science & Engineering
Karunya University, Coimbatore, India

Abstract: Intrusion detection systems are software and/or hardware components that monitor computer systems and analyse events occurring in them for signs of intrusions. Due to widespread diversity and complexity of computer infrastructures, it is difficult to provide a completely secure computer system. Therefore, there are numerous security systems and intrusion detection systems that address different aspects of computer security. This paper first provides taxonomy of IDS with a simple description. Secondly a common architecture of Intrusion detection system (IDS) and their basic characteristics are presented. Thirdly taxonomy of anomaly based intrusion detection system gives brief description along with the properties of advantages and disadvantages.

Keywords: Intrusion Detection System; Cyber Security; Anomaly Based Intrusion Detection System; Misuse Detection and Anomaly Detection;

I. Introduction

The cyber security programs of US organizations do not conflicting the persistence, tactical skills, and technological competence of their potential cyber adversaries. Today, common criminals, organized crime rings, and nation-states leverage sophisticated techniques to launch attacks that are highly targeted and very difficult to detect. Particularly worrisome are attacks by tremendously skilled threat actors that attempt to steal highly sensitive and often very valuable intellectual property, private communications, and other strategic assets and information. (US cybercrime: Rising risks, reduced readiness 2014)[1].

In fact, the US Director of National Intelligence has ranked cybercrime as the top national security threat, higher than that of terrorism, espionage, and weapons of mass destruction. Underscoring the threat, the FBI 2013 notified 3,000 US companies ranging from small banks, major defense contractors, and leading retailers that they had been victims of cyber intrusions.

In today's volatile cybercrime environment, nation-states and other criminals continually and rapidly update their policies to maintain an advantage against advances in security safeguards implemented by businesses and government agencies. Recently, for instance, hackers engineered a new round of distributed denial of service (DDoS) attacks that can generate traffic rated at a staggering 400 gigabits per second, the most powerful DDoS assaults to date.

Critical infrastructure systems used in electrical power distribution, oil and gas pipelines, water supplies, and transportation are particularly vulnerable because their legacy architecture may be easier to compromise. Similarly, the coming year could bring a new wave of strikes on industries that have not migrated critical systems from the Windows XP operating system, which Microsoft no longer supports with security updates. Despite a six-year advance notice that Microsoft would end XP support in April 2014, utility companies continue to run the outdated operating system. Many cash ATMs also use Windows XP, although some employ a simplified embedded version that Microsoft will support until January 2016

PwC's Annual Global CEO Survey 2014 found 69% of US respondents reported they were worried about the impact of cyber threats to their growth prospects, significantly higher than 49% of global CEOs who reported the same unease

One reason for the heightened concern is the high financial costs of cybercrime. PwC's 2014 Global Economic Crime Survey found that 7% of US organizations lost $1 million or more due to cybercrime incidents in 2013, compared with 3% of global organizations; furthermore, 19% of US entities reported financial losses of $50,000 to $1 million, compared with 8% of worldwide respondents [1].

This paper is planned as follows like this, section I gives introduction, section II gives brief description on taxonomy of IDS, section III explains classification of Anomaly based intrusion detection techniques.

II. General Architecture/Framework of Intrusion Detection System

The first model of intrusion detection system was developed by Dorothy E. Denning in 1988[2].This was the basic model of all intrusion detection system although these systems are extremely miscellaneous in the techniques to gather and analyses data, most of them depend on a relatively general architectural framework [3] (Fig.1), which consists of the following components:

Detector: It is also a known as Intrusion Detection (ID) analysis engine. It processes the data Collected from sensors to classify intrusive activities.

Data gathering device: It is also known as sensors. It is responsible for collecting data from the Monitored system.

Knowledge base: It is also known database. It contains information collected by the sensors, but in pre-processed format (e.g. knowledge base of attacks and their signatures, filtered data, data profiles, etc.). This information is usually provided by network and security experts.

Configuration: The device provides information about the current state of the intrusion detection system (IDS).

Response component: This initiates actions when an intrusion is detected. The responses can either be automated (active) or involve human interaction (inactive).

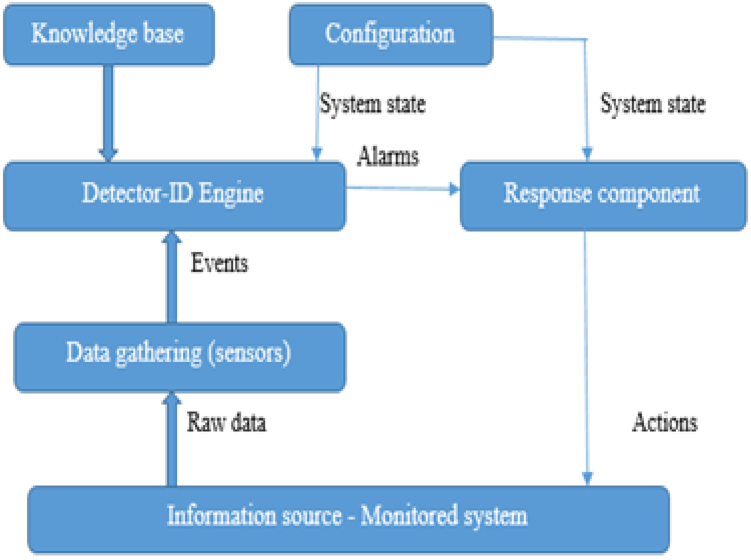

Fig.1 Basic architecture of Intrusion detection System

III. Literature Survey on Taxonomy of Intrusion detection system (IDS)

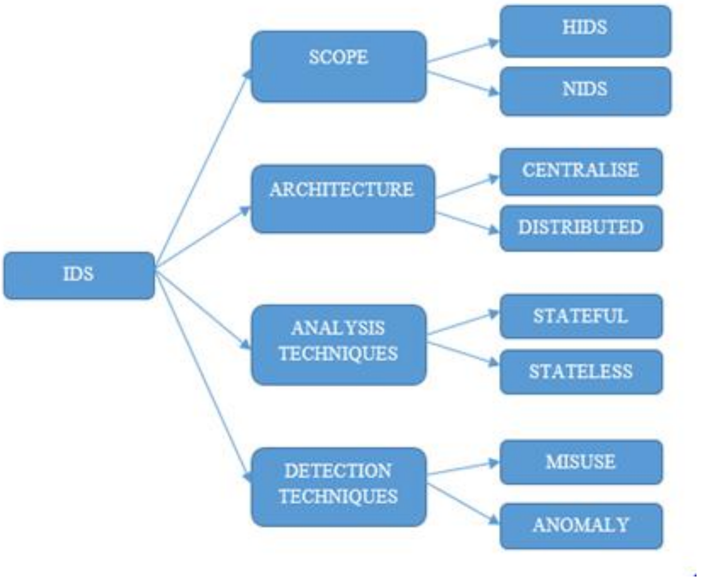

Fig 2: Taxonomy of intrusion detection system

a) Classification on the basis of Architecture

Centralized

- All the operations are performed by the same machine
- It is simpler to realize
- It has single point of failure.

Distributed

- It consists of few elements (a console and sensors) Sensors which generate security events
- The console to monitor events and alerts and control the sensors
- The central Engine that records the events and generate alarms
- Its need is to deal with different data formats.
- It needs secure communication protocol.

b) Classification on the basis of analysis techniques

Stateless

- It treats each event independently of the others.
- It has simple system design.
- It has high processing speed.

State full

- It maintains information about past events
- The effect of a certain event depends on its position in the events stream
- It has more complex system design
- It is more effective in detecting distributed attacks

c) Classification on the basis of detection (approaches) techniques.

Misuse detection: Misuse detection is effective in detecting only known attacks. It recognizes intrusion by looking for patterns of traffic or of application data supposed to be malicious. The main concerns in misuse detection system are how to write a signature that includes all possible variations of the pertinent attack. And how to write signatures that do not also match non-intrusive activity.(Fig 3) shows that, it identifies intrusions by matching monitored events to patterns or signatures of attacks[4]. The attack signatures are the characteristics associated with successful known attacks The main advantage of misuse detection is high accuracy in detecting known attacks. It has limited detection ability on signature database. Unless new attacks are transformed into signatures and added to the database, misuse-based IDS cannot detect any new type of attacks. Different approaches such as expert systems, signature analysis, and state transition analysis are employed in misuse detection

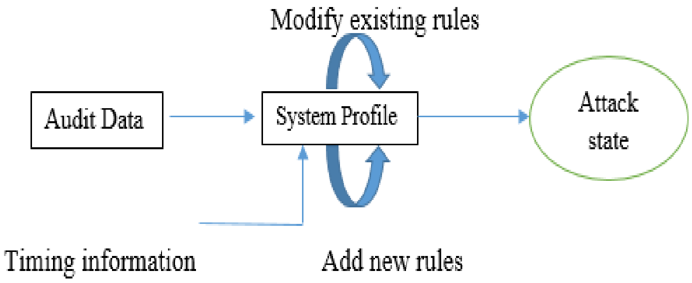

Fig 3: Misuse detection [4]

Anomaly detection: It identifies intrusions by categorizing activity as either anomalous or normal. It needs a training phase to identify normal activity. Anomaly detection technique detects “new” attacks. It generates more false alarms like false positive and false negative than a misuse based IDS [4]. (Fig 4) shows that, Anomaly detection assumes that intrusions are anomalies that necessarily differ from normal performance. It establishes a profile for normal operation and any significantly deviations from the profile as attacks. The major advantage of anomaly detection is it can detect only unknown attacks However, this advantage is paid for in terms of a high false positive rate because, in practice, anomalies are not necessarily intrusive. The number of new attacks increases quickly, it is hard for a misuse detection approach to maintain a high detection rate. In addition, exhibiting attacks is a highly qualified and time- consuming job that leads to a heavy workload of maintaining the signature database .

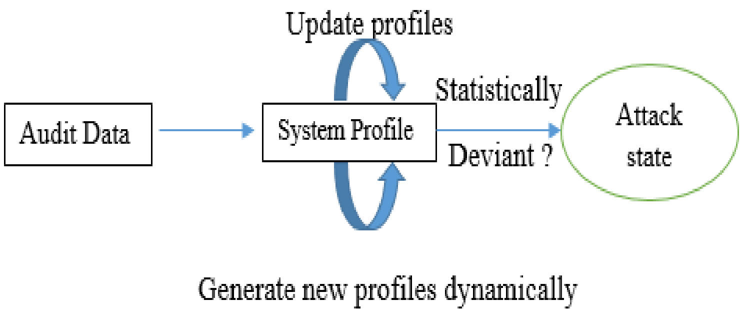

Fig.4 Anomaly detection

d) Classification of Anomaly based intrusion detection system techniques

Anomaly detection is based on a host or network. Many distinct approaches are used based on kind of handling related to behavioral model. Anomaly detection system is mainly classified into four classifications. They are as follows;

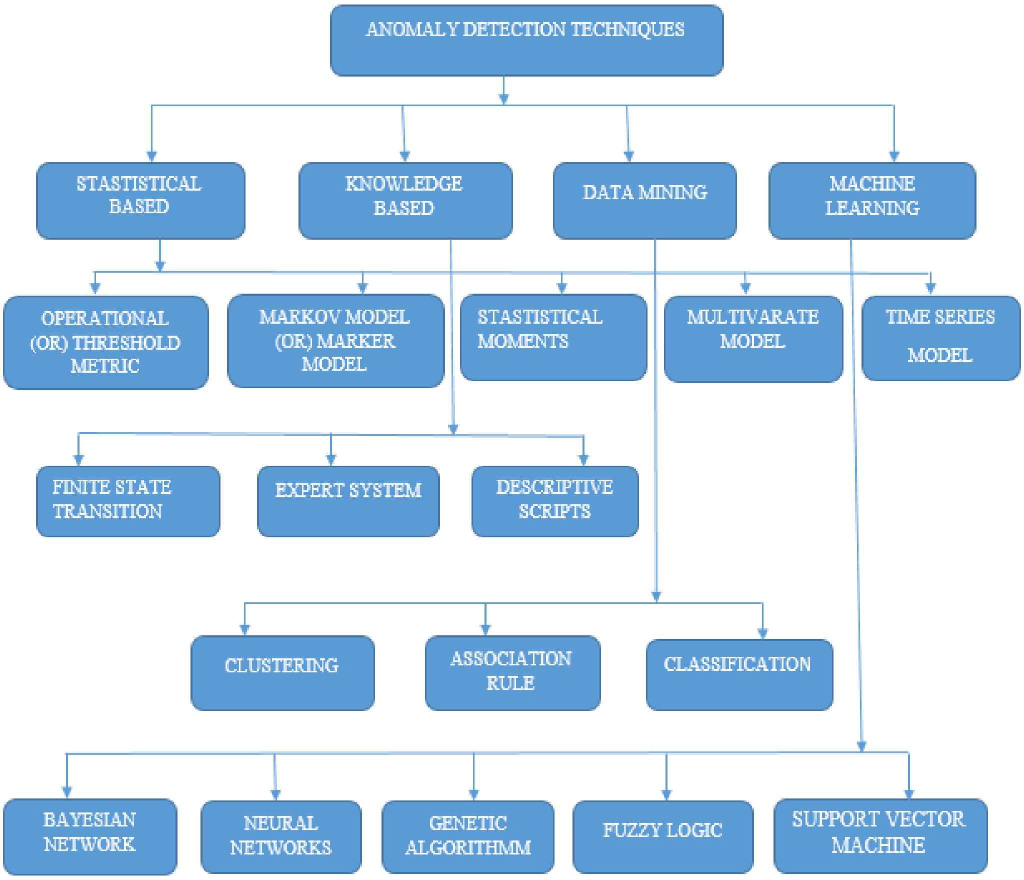

Fig. 5 Classification of Anomaly based intrusion detection system

e) Classification on the basis of Statistical based intrusion detection system

Operational (or) Threshold Metric: Operational Model makes the hypothesis that an anomaly can be recognized through a comparison of an observation with a predefined limit. This model is frequently used in the circumstances where a specific number of events, (i.e., failed logins), is a direct in dictation of a probable attack [3].

Markov Model (or) Marker Model: The Markov Model is used to determine with the event counter metric for the normality of a particular event, based on the events which preceded it. It characterizes each observation as a specific state and utilizes a state transition matrix to determine if the probability of the event is high (normal) based on the preceding events. This model is particularly useful when the sequence of activities is particularly important [5].

Statistical Moments (or) Mean and Standard Deviation: The Average and Standard Deviation Model is based on the traditional statistical determination of the regularity of an observation based on its position relative to a specified confidence range. This model offers the advantage that it “learns” a user’s performance over time instead of requiring previous knowledge of the user’s activities. Finally, this model can establish a basis for the identification of potential anomalies for each user and classify potential problems from users who constantly behave in a manner which would indicate normally, the system resource is misused system. This is mainly useful in identifying what is normal for an individual user without depend on a comparison with other users [3].

Multivariate Model: The Multivariate Model is built upon the Average and Standard Deviation Model. The difference between these two approaches is that the Multivariate Model is based on a correlation of two or more metrics. This model permits the identification of potential anomalies where the complexity of the situation requires the comparison of multiple parameters [3].

Time Series Model: The Time Series Model tries to recognize anomalies by reviewing the order and time interval of activities on the network. If the probability of the occurrence of an observation is low, then the event is considered as abnormal. This model provides the facility to progress over time based on the activities of the users [3].

f) Classification on the basis of Knowledge Based intrusion detection

Finite State Transition: The State transition analysis approach uses state transitions of the system to categorise intrusions. It constructs the state transition diagram, which is the graphical illustration of intrusion performance as a sequence of state variations that lead from an original protected state to a target negotiated state. It lists only the dangerous events that must occur for the real completion of the intrusion. By using of the audit trail as input, an analysis tool can be developed to equal the state changes made by the user to state transition diagrams of known diffusions. State transition diagrams are transcribed to look like the states of a real computer system, and these diagrams form the basis of a rule-based expert system for detecting diffusions, called the state transition analysis tool (STAT) [6]. The advantage is to it detects intrusions self-determining of the audit trial record. It is able to detect difficult attacks, variations to known attacks and attacks covered across multiple user sessions. It has to be used along with some anomaly detector, because USTAT can detect only misuse intrusions. The disadvantages of the system are that it can only constructs patterns from categorizations of events but not from more complex forms and therefore some attacks cannot be detected, as they cannot be exhibited with state transitions [6].

Expert system: An expert system, knowledge about a problem domain is represented by a set of rules. The rules consist of two parts, antecedent, which describes when the rule should be applied and consequent, which describes the action(s) that should be taken if its antecedent is fulfilled. A rule is excited when pattern-matching techniques determine that detected data matches or fulfils the antecedent of a rule. The rules may identify single auditable events that represent significant danger to the system by themselves, or they may recognize a sequence of events that represent an entire diffusion scenario. The disadvantage of the expert system is an intrusion that does not activate a rule will not be detected by the rule based approach. The maintaining and updating a difficult rule-based system can be problematic [6].

Descriptive Scripts: The numerous proposals for scripting languages, which can describe signatures of attacks on computers and networks, are given by the Intrusion Detection community. All of these scripting languages are capable for identifying the sequences of particular events that are revealing of attacks [5].

g) Classification on the basis of data mining based on intrusion detection

Clustering: The Clustering is an unsupervised technique for finding patterns in unlabelled data with many extents (number of attributes).k means clustering is used to find natural groupings of similar instances. The Histories that are far from any of these clusters show unusual activity that may be part of a new attack [7]. The various clustering approaches are Density-based methods, k-mean clustering [8], Mode based methods, Partitioning methods and Hierarchy methods [9].

Association Rule: Association rule based data mining techniques is for anomaly detection where raw data was transformed into ASCII network packet data, which in turn was changed into connection-level information. These connection level records contained connection features like service, duration, etc. Association rules were then applied to this data to create models to detect intrusions.

Classification: An intrusion detection system that classifies audit data as normal or anomalous based on a set of rules, patterns or other associated techniques can be mainly defined as a classification-based intrusion detection system [10]. The classification process usually involves the following steps:

- It identifies class attributes and classes from training data.
- It identifies attributes for classification.
- It learns a model using the training data.
- It uses the learned model to classify the unknown data samples.
- The various classification approaches are Decision trees (DT), ID3, C4.5 [8] and C5.0 [11], are used for classification.

h) Classification on the basis of machine learning based algorithm

Bayesian network: Bayesian network is a graphical representation that translates probabilistic relationships between variables of attention. It is used in the combination with statistical techniques [10]. It has several advantages for data analysis. It converts the interdependencies between the variables, that it can handle circumstances where data is absent. It has the capability to signify new relationships. It is used to predict the consequences of an action. It has both causal and probabilistic relationships, can be used to model problems when there is a need to combine previous knowledge with data. [12] Explains that NBC technique is based on the Bayesian theorem and is particular suited for the dimensionality of the input is high. In spite of its simplicity Naïve Bayes can often outperforms more sophisticated classification method. It works on strong individuality relation assumption, features are independent in the context of a session class and the probability of one attribute does not affect the probability of the other.

Neural networks: Artificial Neural Network (ANN) is one of the most widely used approaches. It has been successful in solving many difficult practical problems. ANN has been effectively used in IDS [13]. The major drawbacks of ANN-based IDS exist in two features; it has lower detection precision, especially for low-frequent attacks and weaker detection stability. The distribution of different kinds of attacks is unfair. For low-frequent attacks, the leaning sample size is too small compared to high-frequent attacks. ANN has difficult to learn the characters of attacks and detection precision is much lower.Neural Networks has strong discrimination and generalisation abilities. It is a model simulation of the neurons in the human brain .Neural Network is collected of several number of highly interconnected processing elements (neurons).Each handling element is essentially a summing element followed by a real function. The production of each neuron is fed as the input to all of the neurons in the following layers [14].

Genetic Algorithm: GA is fundamentally a type of search algorithm used to solve a wide variety of difficulties. Its goal is to create optimal solutions to problem. A potential solution is prearranged as a sequence of bits, characters or numbers. This unit of converting is called a gene, and the converting sequence is known as a chromosome.GA begins with a set of chromosomes, called a population, and a valuation function which measures the fitness value of each chromosome. Usually, an original population of chromosomes is created by complete randomization [15].Chromosomes are weighed by the fitness function. Based on their fitness values, better chromosomes are selected as parents by selection procedure, and then the parents perform crossover and mutation to form new children chromosomes. Finally, some chromosomes in the current generation are replaced by the new ones, if necessary, to form the next generation. The development continues until some predefined situation is met, such as the number of repetitions reached or a satisfactory fitness value appearing. A large number of repetitions is then executed that low execution programs are exchanged by genetic recombination of high execution programs. The program with a low fitness measure is deleted and does not endure for the next computer reiteration

Fuzzy logic: The Fuzzy logic techniques has been used in the area of computer and network security since the late 1990's .It plays an important role in relating anomaly based intrusion detection system. The fuzzy logic part of the system is mostly responsible for both handling the large number of input parameters and dealing with the fuzziness of the input data. When combined with data mining, it reduces the size of the input data sets and selects features that highlight anomalies. Fuzzy logic can be a real means of defining network attacks [6].

Support Vector Machine: The Support Vector Machine (SVM) approach is a machine learning method used for classification. It is a highly well-organized and real in the area of many applications like pattern recognition, image processing, fraud detection, text categorization etc., Because of its accuracy, robustness it provides a best ordering function to distinguish between members of the two classes in the training data. The disadvantage is memory constraint and computational complexity. The various approaches are developed to overcome the limits which are mainly classified into decomposition based and variant based algorithms. It is initially developed to perform binary classification and its classification is very limited. The various techniques are decomposition based method it overcomes memory and limitation, variant based technique is used to reduce the Computational complexity, and the multiclass based methods handle the multi class classification [17].

IV. Conclusion

In this paper various approaches of Intrusion Detection Systems techniques are discussed. The popularity of using Internet contains some risks of network attacks. Intrusion detection is one main research problems in network security; the main aim is to identify unusual access or attacks to secure inside networks. There is no single criterion can be used to completely protect against computer network intrusion. There is no single version of it that can be used as a standard solution against all possible attacks. It is both technically difficult and economically costly to constructs and maintains computer system, networks that are not susceptible to attacks. The technique is to be selected depends on the specifications of the kind of anomalies that the system is imaginary to face, the kind and performance of the data, the environment in which the system is occupied, the cost and computation limitations and the security level mandatory.

References

1. US cybercrime: Rising risks, reduced readiness Key findings from the 2014 US State of Cybercrime Survey www.pwc.com/cybersecurity
2. Denning E. Denning, An intrusion detection model, IEEE Transactions on Software Engineering 1997:2 22–8.
3. Aleksandar Lazarevic, Vipin Kumar, Jaideep Srivastava INTRUSION DETECTION: A SURVEY.
4. Poonam Dabas* and Rashmi Chaudhary, Survey of Network Intrusion Detection Using K-Mean Algorithm, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 3, March 2013.
5. V. Jyothsna, V. V. Rama Prasad and K. Munivara Prasad. A Review of Anomaly based Intrusion Detection Systems, International Journal of Computer Applications (0975 – 8887) Volume 28– No.7, August 2011
6. Sandhya Peddabachigari, Ajith Abraham,Crina Grosan and Johnson Thomas , Modeling intrusion detection system using hybrid intelligent systems, Computer Applications 30 114–132, 2007,ELSEVIER.
7. Manasi Gyanchandani, J.L.Rana and R.N.Yadav, axonomy of Anomaly Based Intrusion Detection System: A Review .International Journal of Scientific and Research Publications, Volume 2, Issue 12, December 2012 ISSN 2250-3153
8. Amuthan Prabakar Muniyandi, R. Rajeswari, R. Rajaram , Network Anomaly Detection by Cascading K-Means Clustering and C4.5 Decision Tree algorithm , In Procedia Engineering 30 , 174 – 182, 2012.ELSEVIER.
9. Shi-Jinn Horng, Ming-Yang Su, Yuan-Hsin Chen, Tzong-Wann Kao, Rong-Jian Chen, Jui-Lin Lai and Citra Dwi Perkasa, A novel intrusion detection system based on hierarchical clustering and support vector machines, Expert Systems with Applications 38 , 306–313, 2011,ELSEVIER.
10. Animesh Patcha*, Jung-Min Park, An overview of anomaly detection techniques: Existing solutions and latest technological trends .Computer Networks 51, 3448–3470, 2007, ELSEVIER.
11. Vahid Golmah, An Efficient Hybrid Intrusion Detection System based on C5.0 and SVM .International Journal of Database Theory and ApplicationVol.7, No.2 (2014), pp.59-70, 2014
12. Neelam Sharma and Dr. Saurabh Mukherjee, A Novel Multi-Classifier Layered Approach to Improve Minority Attack Detection in IDS, In Procedia Technology 6 , 913 – 921, 2012, ELSEVIER.
13. Gang Wang, Jinxing Hao,Jian Ma and Lihua Huang, A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering .Expert Systems with Applications 37, 6225–6232, 2010, ELSEVIER.
14. Parveen Kumar and Nitin Gupta, A Hybrid Intrusion Detection System Using Genetic–Neural Network, International Journal of Engineering Research and application (IJERA) ISSN: 2248-9622 ,National Conference on Advances in Engineering and Technology (2014)

15. Ming-Yang Su, Real-time anomaly detection systems for Denial-of-Service attacks by weighted k-nearest-neighbour classifiers, Expert Systems with Applications 38 , 3492–3498, 2011, ELSEVIER.

16. Dr. Amit Ganatra , Variations of Support Vector Machine classification Technique: A survey, International Journal of Advanced Computer Journal of Advanced Computer Research (IJACR) ,(Vol.2 ,No.6), 2012.

AUTHOR(S) PROFILE

Meesala Shobha Rani, received the B.Tech degree in Computer Science and Engineering from JNTU Ananthapur, and M.Tech degree in Computer Science and Engineering with a specialization of Computer and Communication Engineering from the Karunya University, Coimbatore respectively**.**

S. Basil Xavier received the M.Tech degree in Karunya University. Currently he is working Assistant Professor in Karunya University, Coimbatore respectively.